Datasets and guide: downloads for reconstructing British census responses 1851-1881 for the BBCE

Piero Montebruno

pfm27@cam.ac.uk

Working Paper 19:
Working paper series from ESRC project ES/M010953:
Drivers of Entrepreneurship and Small Businesses
Principal Investigator: Prof. Robert J. Bennett.

University of Cambridge, Department of Geography and Cambridge Group for the History of Population and Social Structure, Downing Place, Cambridge, CB2 3EN, UK.

September 2019

Supplementary material for BBCE data deposit and
WP 9: Bennett, Robert, J., Montebruno, Piero, Smith, Harry, and van Lieshout, Carry (2018) *Reconstructing entrepreneurship and business numbers for censuses 1851-81,* and
WP 9.2: Bennett, Robert, J., Montebruno, Piero, Smith, Harry, and van Lieshout, Carry (2019) *Reconstructing business proprietor responses for censuses 1851-81: a tailored logit cut-off method.*

Comments are welcomed on this paper: contact the author as above.

Keywords: Entrepreneurship, Employers, Self-employment, Small businesses, Census 1851, Census 1861, Census 1871, Census 1881, Reconstruction, England, Wales, Scotland.

JEL Codes: L26, L25, D13, D22

Datasets and guide: downloads for reconstructing British census responses 1851-1881 for the BBCE

Piero Montebruno

Working Paper 19: ESRC project ES/M010953: Drivers of Entrepreneurship and Small Businesses, University of Cambridge

1. Introduction.

This paper introduces the results of the estimation methods used to supplement data where responses to the 1851-1881 censuses were incomplete (in not allowing explicit identification of entrepreneurial status). These are used to reconstruct estimates of the full population of entrepreneurs in England and Wales, and Scotland. The paper provides downloads of the intermediate variables used in the reconstruction, gives their definitions, and provides a detailed guide to understanding and use.

The reconstruction method has been developed to adjust census responses to supplement those individuals who gave explicit responses as to entrepreneur status, with those where entrepreneur status has to be estimated from their characteristics of occupation, demography, etc. These supplemented individuals are used as a reconstructed set of data in the *British Census of Entrepreneurs 1851-1911* (BBCE), deposited at UKDA. This paper derives from the ESRC-supported project ES/M010953 *Drivers of Entrepreneurship and Small Businesses*. The data referred to are derived from the electronic census data made available through the database deposit of the original CEBs at the UKDA: *The Integrated Census Microdata* (I-CeM). The I-CeM version used derives from version 2 of these data.[1] An overview of the project is provided in WP 1, WP 2 defines in detail the different censuses

[1] Higgs, Edward and Schürer, Kevin (University of Essex) (2014) *The Integrated Census Microdata (I-CeM)* UKDA, SN-7481; K. Schürer, E. Higgs, A.M. Reid, E.M Garrett, *Integrated Census Microdata, 1851-1911, version V. 2 (I-CeM.2)*, (2016) [data collection] UK Data Service SN: 7481.

used for this working paper and the challenges they present for identifying entrepreneurs over 1851-81. WP 9 and WP 9.2 discuss the reconstruction methods in detail. A full list of all Working Papers is given at the end of this paper.

The reconstruction methods are described in WP 9 for England and Wales as an aggregate estimate (NUM) which we term 'intelligence-led', and as an alternative using a tailored logit cut-off method (IND) in WP 9.2 which is used for both England and Wales, and Scotland.

Note that the data in I-CeM for England and Wales do not currently contain anything for 1871, and if this is added from the FMP records there will be no occupational codes as this was not transcribed. Hence 1871 cannot be used in the reconstructions. BBCE does contain data on 1871 for those employers and own account identifiable directly in the census (Groups 1-6). This is derived from a different source of transcripts (S&N). But these data do not cover the whole 1871 population and so reconstruction of employment status for non-respondents in the census is not possible.

The proprietors identified by both the intelligence-led and tailored cut-offs are available in the UKDA deposit of BBCE by I-CeM RecID for each individual with the variable names:

- EMPSTATUS_NUM (for the outputs derived from WP 9), for both England and Wales, and Scotland.
- EMPSTATUS_IND (for the outputs derived from this WP 9.2), for England and Wales, and

These reconstruction methods are used to generate an employment status as Worker (W), Employer (E), and Own account (OA). These are defined for each of the economically active individuals in Censuses 1851, 1861, and 1881 for England and Wales, and in Censuses 1851, 1861, 1871 and 1881 for Scotland:

1. Method NUM. This method, used only for England and Wales, is guided with the objective of providing the best possible estimate of the number of Workers (W), Employers (E), and Own account proprietors (OA) in each of occupational categories. This is for a supplemented set of Occodes to those defined in I-CeM which allow finer detail of occupational descriptors to be used to separate W, E, and OA. These are referred to as 882 SubOccodes. Because this method estimates the aggregate numbers of W, E, and OA for each type of the 882 SubOccodes (and not the estimates for

individuals), the individual employment status may result from a random allocation. This method is more exact at an aggregate level but unsatisfactory at an individual level. More detailed discussion of the method is provided in the WP 9.

2. Method IND. This uses the same classical Logit Model as for Method NUM in two stages, one for entrepreneurs, that is W and Entrepreneurs, and subsequently one to separate Entrepreneurs between E and OA. This method differs from Method NUM by calculating a score for each individual using all the available information in the censuses, but with variable logit cut-offs. This has no need to use any kind of random assignment. Hence, this is the preferable estimate for identifying the employment status of each individual in the British censuses. But it is less accurate for estimating aggregate categories. More detailed discussion is provided in WP 9.2, which also compares the outputs form the two methods.

For both reconstruction methods the farm entrepreneurs are treated in a separate way using a farm model. This is possible because the census questions collected additional data for farmers on the acreage they occupied. The farm responses were also more complete in the census, in general, than non-farmers. This allows a more direct method of estimating farm status between W, E and OA without the need for the more general reconstruction methods. The estimates of farmer status are included in the downloads with this paper. Hence, the downloads are complete for all occupational categories, but farmers derive from the separate farm model. The details of the farm model are given in WP 9 and discussed more fully in Montebruno et al. (2019).

The detailed dataset with all the variables and steps for each reconstruction method are given in separate files available for download linked to this WP. The reconstruction method allows an interested researcher to access the full business population of economically active individuals, in particular, it permits access to the entrepreneurial population separated between employers and own account proprietors. This effort creates for the first time a full database on economically active individuals and entrepreneurs using data from I-CeM linked to the BBCE. The paper provides detailed documentation of how the data base was created and can be used as a short tutorial guide on its realisation.

The paper in the next section presents the tutorial guide based on the method NUM reconstruction. Then Section 3 lists the variables of the Reconstruction database for England and Wales (both NUM and IND methods) and for Scotland (IND method).

2. The NUM-method reconstruction: a brief tutorial.

2.1. Revisiting NUM-method reconstruction's outputs.

This tutorial guide shows how to retrieve any wanted data from the NUM-method reconstruction dataset. First, it shows how to "clean" the data to produce tables providing expected reconstruction figures. Take for instance, 1881. Being ReadyTypeE_81 (R) (see below for a comprehensive list of all the variables and their definition in the accompanying database) the final allocation for the employers reconstruction and New Employ Code (N) the last allocation for the overall reconstruction, one should expect that:

1. No workers where ReadyType_81 is different from missing and all the workers missing
2. All the employers in the cell (R=1, N=2) with 0 in (R=0, N=2) and (R=., N=2). (Thus, a column of only zeros but for the cell that represents employers in both variables)
3. Like own accounts but with (R=0, N=3) concentrating all the own accounts and zeros in the two cells below.

But as the following tabulation shows, this is not the case.

```
. tab ReadyTypeE_81 NewEmployCode ,m

ReadyTypeE |          NewEmployCode
       _81 |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 |     1,301     11,524    902,220 |   915,045
         1 |    11,887    515,268     37,958 |   565,113
         . | 9,503,555      9,060     13,238 | 9,525,853
-----------+---------------------------------+----------
     Total | 9,516,743    535,852    953,416 |11,006,011
```

The first thing that confuses the relationships is the only-workers. In the employer reconstruction some only-worker categories were treated as normal and there were some employers and own accounts allocations that needed to be reallocated to workers at the end.

So if you take out the only worker allocations the first column gets better and become clear as wanted.

```
. tab ReadyTypeE_81 New if Other!=-1,m

ReadyTypeE |          NewEmployCode
       _81 |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 |         0     11,524    902,220 |   913,744
         1 |         0    515,268     37,958 |   553,226
         . | 5,370,365      9,060     13,238 | 5,392,663
-----------+---------------------------------+----------
     Total | 5,370,365    535,852    953,416 | 6,859,633
```

As noted in introduction, the farmers had an independent allocation (see WP 9; Montebruno et al., 2019). This also confuses the allocations. If you take out the farmers (Farmer_sample=0), then the cells (R=0, N=2) and (R=1, N=3) become zero which is a further correction and we are almost with the wanted results.

```
. tab ReadyTypeE_81 New if Farmer_s==0  & Other!=-1,m

ReadyTypeE |          NewEmployCode
       _81 |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 |         0          0    879,550 |   879,550
         1 |         0    399,994          0 |   399,994
         . | 5,334,506      8,999     13,146 | 5,356,651
-----------+---------------------------------+----------
     Total | 5,334,506    408,993    892,696 | 6,636,195
```

Finally, those variables in I-CeM which are assigned a value of U have to be controlled for; these are where data are incomplete (e.g. gender is not given). The Us were allocated according to the proportions of W, E, OA in the sample, which is different from the employers reconstruction. So take out the Us and we are done!

```
. tab ReadyTypeE_81 New if Farmer_s==0 & Us==0 & Other!=-1,m

ReadyTypeE |          NewEmployCode
       _81 |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 |         0          0    879,550 |   879,550
         1 |         0    399,994          0 |   399,994
         . | 5,267,610          0          0 | 5,267,610
-----------+---------------------------------+----------
     Total | 5,267,610    399,994    879,550 | 6,547,154
```

So in a nutshell, different reconstructions techniques for workers-only, farmers, and Us are the cause of the unclear tabulation between the final allocation variable of the employer reconstruction (R) and the final allocation variable of the overall reconstruction (N). So if you want to be safe USE ALWAYS NewEmployCode (N), and if you want to see the details use the final allocation of the employer reconstruction (R) but after removing the workers only, the farmers, and the Us.

A similar argument can explain why the Entrepreneur reconstruction final allocation variable ReadyType81CORREGIDA and the overall final allocation variable do not coincide, but their strict relation becomes apparent if one excludes workers only, farmers, and Us as the following results show.

```
. tab ReadyType81CORREGIDA NewEmployCode ,m

ReadyType8 |          NewEmployCode
1CORREGIDA |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 | 9,416,207          0          0 | 9,416,207
         1 |    66,078    534,839    950,632 | 1,551,549
         . |    34,458      1,013      2,784 |    38,255
-----------+---------------------------------+----------
     Total | 9,516,743    535,852    953,416 |11,006,011


. tab ReadyType81CORREGIDA NewEmployCode if Others!=-1,m

ReadyType8 |          NewEmployCode
1CORREGIDA |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 | 5,303,469          0          0 | 5,303,469
         1 |    51,320    534,839    950,632 | 1,536,791
         . |    15,576      1,013      2,784 |    19,373
-----------+---------------------------------+----------
     Total | 5,370,365    535,852    953,416 | 6,859,633


. tab ReadyType81CORREGIDA NewEmployCode if Farmer_s==0 & Others!=-
1,m

ReadyType8 |          NewEmployCode
1CORREGIDA |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 | 5,267,610          0          0 | 5,267,610
         1 |    51,320    407,980    889,937 | 1,349,237
         . |    15,576      1,013      2,759 |    19,348
-----------+---------------------------------+----------
     Total | 5,334,506    408,993    892,696 | 6,636,195
```

```
. tab ReadyType81CORREGIDA NewEmployCode if Farmer_s==0 & Us==0 &
Others!=-1,m

ReadyType8 |          NewEmployCode
1CORREGIDA |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 | 5,267,610          0          0 | 5,267,610
         1 |         0    399,994    879,550 | 1,279,544
-----------+---------------------------------+----------
     Total | 5,267,610    399,994    879,550 | 6,547,154 -----------
---------------------------------------
```

Same for 1861

```
. tab ReadyTypeE_61 NewEmployCode ,m

ReadyTypeE |          NewEmployCode
       _61 |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 |    19,430     13,158    675,710 |   708,298
         1 |    61,694    438,440     26,613 |   526,747
         . | 6,985,980      3,107      7,030 | 6,996,117
-----------+---------------------------------+----------
     Total | 7,067,104    454,705    709,353 | 8,231,162


. tab ReadyTypeE_61 New if Other!=-1,m

ReadyTypeE |          NewEmployCode
       _61 |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 |    12,177     13,158    675,710 |   701,045
         1 |    23,083    438,440     26,613 |   488,136
         . | 3,900,150      3,107      7,030 | 3,910,287
-----------+---------------------------------+----------
     Total | 3,935,410    454,705    709,353 | 5,099,468


. tab ReadyTypeE_61 New if Farmer_s==0  & Other!=-1,m

ReadyTypeE |          NewEmployCode
       _61 |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 |         0          0    661,618 |   661,618
         1 |         0    307,955          0 |   307,955
         . | 3,896,736      2,818      6,639 | 3,906,193
-----------+---------------------------------+----------
     Total | 3,896,736    310,773    668,257 | 4,875,766
```

```
. tab ReadyTypeE_61 New if Farmer_s==0 & Us==0 & Other!=-1,m

ReadyTypeE |          NewEmployCode
       _61 |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 |         0          0    661,618 |   661,618
         1 |         0    307,955          0 |   307,955
         . | 3,853,783          0          0 | 3,853,783
-----------+---------------------------------+----------
     Total | 3,853,783    307,955    661,618 | 4,823,356



.
. tab RReadyType NewEmployCode ,m

   ESTE SI |
ReadyType6 |          NewEmployCode
         1 |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 | 6,890,909          0        211 | 6,891,120
         1 |    81,272    451,913    702,526 | 1,235,711
         . |    94,923      2,792      6,616 |   104,331
-----------+---------------------------------+----------
     Total | 7,067,104    454,705    709,353 | 8,231,162


. tab RReadyType NewEmployCode if Others!=-1,m

   ESTE SI |
ReadyType6 |          NewEmployCode
         1 |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 | 3,856,832          0        211 | 3,857,043
         1 |    35,384    451,913    702,526 | 1,189,823
         . |    43,194      2,792      6,616 |    52,602
-----------+---------------------------------+----------
     Total | 3,935,410    454,705    709,353 | 5,099,468


. tab RReadyType NewEmployCode if Farmer_s==0 & Others!=-1,m

   ESTE SI |
ReadyType6 |          NewEmployCode
         1 |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 | 3,853,783          0          0 | 3,853,783
         1 |       124    308,028    661,671 |   969,823
         . |    42,829      2,745      6,586 |    52,160
-----------+---------------------------------+----------
     Total | 3,896,736    310,773    668,257 | 4,875,766
```

```
. tab RReadyType NewEmployCode if Farmer_s==0 & Us==0 & Others!=-1,m

   ESTE SI |
ReadyType6 |          NewEmployCode
         1 |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 | 3,853,783          0          0 | 3,853,783
         1 |         0    307,955    661,618 |   969,573
-----------+---------------------------------+----------
     Total | 3,853,783    307,955    661,618 | 4,823,356
```

And for 1851

```
. tab ReadyTypeE_51 NewEmployCode ,m

ReadyTypeE |          NewEmployCode
       _51 |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 |    20,602     19,015    642,268 |   681,885
         1 |    64,045    419,764     28,371 |   512,180
         . | 6,336,230     22,154     20,038 | 6,378,422
-----------+---------------------------------+----------
     Total | 6,420,877    460,933    690,677 | 7,572,487


. tab ReadyTypeE_51 New if Other!=-1,m

ReadyTypeE |          NewEmployCode
       _51 |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 |    15,610     19,015    642,268 |   676,893
         1 |    21,972    419,763     28,371 |   470,106
         . | 3,330,139     22,153     20,038 | 3,372,330
-----------+---------------------------------+----------
     Total | 3,367,721    460,931    690,677 | 4,519,329


. tab ReadyTypeE_51 New if Farmer_s==0  & Other!=-1,m

ReadyTypeE |          NewEmployCode
       _51 |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 |     4,293      5,697    628,105 |   638,095
         1 |         0    278,484          0 |   278,484
         . | 3,326,557     21,036     17,912 | 3,365,505
-----------+---------------------------------+----------
     Total | 3,330,850    305,217    646,017 | 4,282,084
```

```
. tab ReadyTypeE_51 New if Farmer_s==0 & Us==0 & Other!=-1,m

ReadyTypeE |          NewEmployCode
       _51 |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 |     4,293      5,697    628,105 |   638,095
         1 |         0    278,484          0 |   278,484
         . | 3,268,874     12,486      5,570 | 3,286,930
-----------+---------------------------------+----------
     Total | 3,273,167    296,667    633,675 | 4,203,509
.

. tab ReadyType51CORREGIDA NewEmployCode ,m

ReadyType5 |          NewEmployCode
1CORREGIDA |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 | 6,197,789     13,195      7,410 | 6,218,394
         1 |    89,094    439,922    672,131 | 1,201,147
         . |   133,994      7,816     11,136 |   152,946
-----------+---------------------------------+----------
     Total | 6,420,877    460,933    690,677 | 7,572,487


. tab ReadyType51CORREGIDA NewEmployCode if Others!=-1,m

ReadyType5 |          NewEmployCode
1CORREGIDA |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 | 3,271,995     13,194      7,410 | 3,292,599
         1 |    42,028    439,921    672,131 | 1,154,080
         . |    53,698      7,816     11,136 |    72,650
-----------+---------------------------------+----------
     Total | 3,367,721    460,931    690,677 | 4,519,329


. tab ReadyType51CORREGIDA NewEmployCode if Farmer_s==0 & Others!=-
1,m

ReadyType5 |          NewEmployCode
1CORREGIDA |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 | 3,268,874     12,486      5,570 | 3,286,930
         1 |     8,739    285,220    629,526 |   923,485
         . |    53,237      7,511     10,921 |    71,669
-----------+---------------------------------+----------
     Total | 3,330,850    305,217    646,017 | 4,282,084
```

```
. tab ReadyType51CORREGIDA NewEmployCode if Farmer_s==0 & Us==0 &
Others!=-1,m

ReadyType5 |          NewEmployCode
1CORREGIDA |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 | 3,268,874     12,486      5,570 | 3,286,930
         1 |     4,293    284,181    628,105 |   916,579
-----------+---------------------------------+----------
     Total | 3,273,167    296,667    633,675 | 4,203,509


========================================
```

And again 1851 after Group replaced Types because of newer extraction

```
=========

. tab ReadyGroupE_51 NewEmployCode ,m

ReadyGroup |          NewEmployCode
      E_51 |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 |    20,643     12,515    647,838 |   680,996
         1 |    64,326    442,225     28,371 |   534,922
         . | 6,335,908      6,193     14,468 | 6,356,569
-----------+---------------------------------+----------
     Total | 6,420,877    460,933    690,677 | 7,572,487


. tab ReadyGroupE_51 New if Other!=-1,m

ReadyGroup |          NewEmployCode
      E_51 |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 |    15,660     12,515    647,838 |   676,013
         1 |    21,972    442,223     28,371 |   492,566
         . | 3,330,089      6,193     14,468 | 3,350,750
-----------+---------------------------------+----------
     Total | 3,367,721    460,931    690,677 | 4,519,329


. tab ReadyGroupE_51 New if Farmer_s==0  & Other!=-1,m

ReadyGroup |          NewEmployCode
      E_51 |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 |     4,343          0    633,675 |   638,018
         1 |         0    299,839          0 |   299,839
         . | 3,326,507      5,378     12,342 | 3,344,227
-----------+---------------------------------+----------
     Total | 3,330,850    305,217    646,017 | 4,282,084
```

```
. tab ReadyGroupE_51 New if Farmer_s==0 & Us==0 & Other!=-1,m

ReadyGroup |          NewEmployCode
      E_51 |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 |     4,343          0    633,675 |   638,018
         1 |         0    296,667          0 |   296,667
         . | 3,268,824          0          0 | 3,268,824
-----------+---------------------------------+----------
     Total | 3,273,167    296,667    633,675 | 4,203,509


.
. tab ReadyGroup51CORREGIDA NewEmployCode ,m

ReadyGroup |
51CORREGID |          NewEmployCode
         A |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 | 6,197,207        483      1,748 | 6,199,438
         1 |    89,765    455,760    678,473 | 1,223,998
         . |   133,905      4,690     10,456 |   149,051
-----------+---------------------------------+----------
     Total | 6,420,877    460,933    690,677 | 7,572,487


. tab ReadyGroup51CORREGIDA NewEmployCode if Others!=-1,m

ReadyGroup |
51CORREGID |          NewEmployCode
         A |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 | 3,271,945        483      1,748 | 3,274,176
         1 |    42,132    455,758    678,473 | 1,176,363
         . |    53,644      4,690     10,456 |    68,790
-----------+---------------------------------+----------
     Total | 3,367,721    460,931    690,677 | 4,519,329


. tab ReadyGroup51CORREGIDA NewEmployCode if Farmer_s==0 & Others!=-
1,m

ReadyGroup |
51CORREGID |          NewEmployCode
         A |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 | 3,268,824          0          0 | 3,268,824
         1 |     8,843    300,748    635,753 |   945,344
         . |    53,183      4,469     10,264 |    67,916
-----------+---------------------------------+----------
     Total | 3,330,850    305,217    646,017 | 4,282,084
```

```
. tab ReadyGroup51CORREGIDA NewEmployCode if Farmer_s==0 & Us==0 &
Others!=-1,m

ReadyGroup |
51CORREGID |          NewEmployCode
         A |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 | 3,268,824          0          0 | 3,268,824
         1 |     4,343    296,667    633,675 |   934,685
-----------+---------------------------------+----------
     Total | 3,273,167    296,667    633,675 | 4,203,509
```

Notice that there are 4,343 misplaced observations, which is explained by the recoding form OAs to Ws compensating for the addition of 4,878 Type 3s (OAs). These observations are kept with their coding to individualise them.

2.2. Final NUM-method reconstruction results.

The following are the final aggregate figures for the critical variable NewEmployCode. That is, the following tables give the number of Workers (1), Employers (2), and Own-account (3) for each of the Censuses years 1851, 1861, and 1881.

```
FINAL NEWEMPLOYCODE TABULATIONS
----------------------------------------
1851

NewEmployCo |
         de |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |  6,420,877       84.79       84.79
          2 |    460,933        6.09       90.88
          3 |    690,677        9.12      100.00
------------+-----------------------------------
      Total |  7,572,487      100.00

1861


NewEmployCo |
         de |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |  7,067,104       85.86       85.86
          2 |    454,705        5.52       91.38
          3 |    709,353        8.62      100.00
------------+-----------------------------------
      Total |  8,231,162      100.00
```

1881

NewEmployCode	Freq.	Percent	Cum.
1	9,516,743	86.47	86.47
2	535,852	4.87	91.34
3	953,416	8.66	100.00
Total	11,006,011	100.00	

2.3. Examples of decisions

The following are two examples of decisions to illustrate how the final employment status is reached and how it is recorded at each step towards the decisions for the NUM-method of reconstruction (the IND-method process is analogous and is also fully recorded in the respective files of the dataset; see later below):

2.3.1. 1851 newRecID of "32". This individual is coded as Ent because of 1901 rounded regression ("Entrepreneur51_01_pr_rou" has value "1", "short" has value "2", and "Decisions" has value "4" or "Use 1901 rounded type extrapolation". A valid annotation here is that some of the recording is redundant to check for consistency and accuracy: here "short" equal to "2" is a redundancy test, also the "Decisions" equal to "4" only repeats that we are using the 1901 rounded regression as a classification tool). In any case, this individual will also be counted as an Ent just because having a "1" in "ReadyType51CORREGIDA", that is before Types where recoded with the new extraction Groups and in "ReadyGroup51CORREGIDA", that is after Groups replaced Types. Thus, our first conclusion is that the individual is an Ent from two independent sources, one of them including extrapolating and predicting for 1851 from a Logistic Regression of 1901 rounded regression. Rounded means that each individual in the SubOccode has the probability rounded and added together and this aggregate rounded probability is what is being chosen. Then it is time to decide if it is an E or an OA among the Ent status. The answer comes, again, from a 1901 Logistic Regression (notice that this a different regression from the Ent one, which has as dependent variable E not Ent this time): "1" or a E (redundancy "shortE" equal to "2"). But this is not the end of the story. The decision is "5" ("Use 1891 ratio

extrapolation") and that means, in this case, that in the aggregate there are too many E in this SubOccode (872), thus the NUM-method, which prioritises the aggregate guess, reallocates the excess of E into OA: calculated as an aggregate figure as individual allocation was by random reassignment of employment status for this SubOccode to match the "1891 ratio extrapolation". This swap in the entrepreneurial code from E to OA is recorded under the variable "RandAllEnt51" which has a "1" meaning that in this case the individual was subject to a random allocation from/to E and OA, and the complementary value "1" in the variable "RandAllEtoOA51", which establishes that the direction of the random allocation is from E to OA in census 1851. Finally, the variable "Employer_sample" which includes all Ent, i.e. E + OA, is also a redundancy and confirms this individual is an Ent. For all the previous analysis, it is not a surprise that the value in the most important variable of the dataset, "NewEmployCode", is "3" or an OA and, in this case, this is the same outcome of the other complementary last values:
"OLD_NewEmployCode", "ALMOSTFINALNewEmployCode", and "RECONSTRUCTION_NewEmployCodE", which does not add much additional information to this case.

2.3.2. 1851 newRecID of "16231412". This individual is in the "Farmer_sample" so all the information in the Ent and E/OA variables is irrelevant as the farm model has been used to employment estimate status. Only the farmer's variables are of interest. The "Type" and the updated "Group" are both "5", i.e. is a farmer OA (who gives acres but not total employees). The farmer's model suggests a random reallocation from "5", OA, to "1", E (" TypeLogit_025" and "FarmerModel5to1025" have value "1") is because the acreages of this farmer are big enough to suggest that the farmer indeed hires labourers but that this information is missing from the census according to the cut-off calculated for this continuous parish in the farmer's model. In this case the final variables are indeed informative. "RECONSTRUCTION_NewEmployCode" has a "3", i.e. the general reconstruction model decision is OA, but as said this individual is a farmer so the farmer's model prevails. Thus, the final allocation is "2" E, the outcome of the farmer's model, and not "3" or OA, the independent outcome of the general reconstruction model. This is confirmed by the variable "NewEmployCode" which has value "2" or E.

3. Variables in the Reconstruction Database.

The following is a comprehensive list of each of the reconstruction variables found in the accompanying "Disaggregated reconstruction database for the British Censuses" (Montebruno, 2019) with a short description of them. The reconstruction process is thus documented in its full realisation with each of the steps clearly named and with each decision recorded in the accompanying database (Montebruno, 2019).

3.1. NUM-method reconstruction variables:

3.1.1. 1851

3.1.1.1. ENT 51

newRecID
Rec ID for 1851

Ent_corr51_91_pr_rou
prediction of being an entrepreneur (0=worker / 1=entrepreneur) by 1891 rounded regression

Entrepreneur51_01_pr_rou
prediction of being an entrepreneur (0=worker / 1=entrepreneur) by 1901 rounded regression

short
Supply of entrepreneurs by 1 (1891 rounded regressions) or 2 (1901 rounded regressions)

RandAllEnt51
If newRecID has been randomly allocated from/to entrepreneur or worker

RandAllWtoEnt51
If newRecID has been randomly allocated from worker to entrepreneur

RandAllEnttoW51
If newRecID has been randomly allocated from entrepreneur to worker

Decisions
Which extrapolation has been decided: 1.Use 1891 unrounded type extrapolation 2.Use 1891 rounded type extrapolation 3.Use 1901 unrounded type extrapolation 4.Use 1901 rounded type extrapolation 5.Use 1891 ratio extrapolation

TypeAlloc51
If newRecID recoded from worker (0) to entrepreneur (1) by imposition of Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers)

ReadyType51CORREGIDA
Final entrepreneur code (before Types where recoded with the new extraction Groups)

Type
Type as extracted from I-CeM

GroupAlloc51
If newRecID recoded from worker (0) to entrepreneur (1) by imposition of Group (Groups 1,3,5 for non-farmers and Groups 1,2,5 for farmers)

ReadyGroup51CORREGIDA
Final entrepreneur code (after Groups replaced Types)

Group
Type as the final extraction from I-CeM (Type=First, Group=Last)

SubOccode
882 Sub-Occodes of economic activities

3.1.1.2. E 51

newRecID
Rec ID for 1851

E_corr51_91_pr_rou
prediction of being an employer (0=own account / 1=employer) by 1891 rounded regression

E_51_01_pr_rou
prediction of being an employer (0=own account / 1=employer) by 1901 rounded regression

shortE
Supply of employers by 1 (1891 rounded regressions) or 2 (1901 rounded regressions)

RandAllEnt51
If newRecID has been randomly allocated from/to employer or own account

RandAllOAtoE51
If newRecID has been randomly allocated from own account to employer

RandAllEtoOA51
If newRecID has been randomly allocated from employer to own account

Decision
Which extrapolation has been decided: 1.Use 1891 unrounded type extrapolation 2.Use 1891 rounded type extrapolation 3.Use 1901 unrounded type extrapolation 4.Use 1901 rounded type extrapolation 5.Use 1891 ratio extrapolation

TypeAllocE_51
If newRecID recoded from own account (0) to employer (1) by imposition of Type (Type 1 means employer)

ReadyTypeE_51
Final employer code (before Type was replaced by the newer extraction Group)

GroupAllocE_51
If newRecID recoded from own account (0) to employer (1) by imposition of Group (Group 1 means employer) (After Groups replaced Types, newer extraction from I-CeM)

ReadyGroupE_51
Final employer code (after Group replaced Type)

SubOccode
882 Sub-Occodes of economic activities

3.1.1.3. FARMERS 51

newRecID
Rec ID for 1851

Type_Farmers
Types of Farmers (slight changes from previous allocation, cfr Type)

TypeLogit_025
Type after the farming model with a probability of 0.25 (odds ratio of 3)

TypeLogit_025NPort4_CAXTON
Type after New Portfolio and NewMarket, Linton, Holywell and Caxton and Cambridge new extractions.

ReadyTypeLogit_025_4to5
Type after the farming model reallocation types 4 to 5

ReadyTypeLogit_025_5to1
Type after the farming model reallocation types 5 to 1

ReadyTypeLogit_025_4to1
Type after the farming model reallocation types 4 to 1

ReadyTypeLogitFinal
Final allocation after new evidence on limited reallocations (marginal adjustments)

FarmerModel5to1025
If farmer model predicted a reallocation from own account to employer (5 to 1) with a probability of 0.25 (odds ratio of 3)

RandAll4to5
If farmers is randomly allocated from types 4 to 5 (to match overall numbers)

RandAll5to1
If farmers is randomly allocated from types 5 to 1 (to match overall numbers)

RandAll4to1
If farmers is randomly allocated from types 4 to 1 (to match overall numbers)

3.1.1.4.

OLD_NewEmployCode Variable New Employ Code before adding the 50,000 newly extracted group.
ALMOSTFINALNewEmployCode Variable New Employ Code before adding 4,800 Type 3 (OAs) and recoding a similar amount from OAs to Ws.
RECONSTRUCTION_NewEmployCodE variable New Employ Code using farmer reconstruction (non-optimised farmers' model)

3.1.2. 1861

3.1.2.1. ENT 61
newRecID
Rec ID for 1861

Ent_corr61_91_pr_rou
prediction of being an entrepreneur (0=worker / 1=entrepreneur) by 1891 rounded regression

Entrepreneur61_01_pr_rou
prediction of being an entrepreneur (0=worker / 1=entrepreneur) by 1901 rounded regression

short
Supply of entrepreneurs by 1 (1891 rounded regressions) or 2 (1901 rounded regressions)

RandAllEnt61
If newRecID has been randomly allocated from/to entrepreneur or worker

RandAllWtoEnt61
If newRecID has been randomly allocated from worker to entrepreneur

RandAllEnttoW61
If newRecID has been randomly allocated from entrepreneur to worker

DecisionB_H
Which extrapolation has been decided: 1.Use 1891 unrounded type extrapolation 2.Use 1891 rounded type extrapolation 3.Use 1901 unrounded type extrapolation 4.Use 1901 rounded type extrapolation 5.Use 1891 ratio extrapolation

TypeAlloc61
If newRecID recoded from worker (0) to entrepreneur (1) by imposition of Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers)

RReadyType
Final entrepreneur code

Type
Type as extracted from I-CeM
SubOccode
882 Sub-Occodes of economic activities

3.1.2.2. E 61

newRecID
Rec ID for 1861

E_corr61_91_pr_rou
prediction of being an employer (0=own account / 1=employer) by 1891 rounded regression

E_61_01_pr_rou
prediction of being an employer (0=own accout / 1=employer) by 1901 rounded regression

shortE
Supply of employers by 1 (1891 rounded regressions) or 2 (1901 rounded regressions)

RandAllEnt61
If newRecID has been randomly allocated from/to employer or own account

RandAllOAtoE61
If newRecID has been randomly allocated from own account to employer

RandAllEtoOA61
If newRecID has been randomly allocated from employer to own account

decision_afternewdecision
Which extrapolation has been decided: 1.Use 1891 unrounded type extrapolation 2.Use 1891 rounded type extrapolation 3.Use 1901 unrounded type extrapolation 4.Use 1901 rounded type extrapolation 5.Use 1891 ratio extrapolation

TypeAllocE_61
If newRecID recoded from own account (0) to employer (1) by imposition of Type (Type 1 means employer)

ReadyTypeE_61
Final employer code

TypeE_61
Type as extracted from I-CeM

SubOccode
882 Sub-Occodes of economic activities

3.1.2.3. FARMER 61

TypeLogit_025NPort4
Final allocation of types for farmers after applying a reallocation of 5s to 1s with a logit farming model using a probability of 0.25 (odds ratio of 3)

3.1.2.4.

RECONSTRUCTION_NewEmployCodE variable New Employ Code using farmer reconstruction (non-optimised farmers' model)

3.1.3. 1881

3.1.3.1. ENT 81

newRecID
Rec ID for 1881

Ent_corr81_91_pr_rou
prediction of being an entrepreneur (0=worker / 1=entrepreneur) by 1891 rounded regression

Entrepreneur81_01_pr_rou
prediction of being an entrepreneur (0=worker / 1=entrepreneur) by 1901 rounded regression

short
Supply of entrepreneurs by 1 (1891 rounded regressions) or 2 (1901 rounded regressions)

RandAllEnt81
If newRecID has been randomly allocated from/to entrepreneur or worker

RandAllWtoEnt81
If newRecID has been randomly allocated from worker to entrepreneur

RandAllEnttoW81
If newRecID has been randomly allocated from entrepreneur to worker

DecisionFinal
Which extrapolation has been decided:
1 - use unrounded 1891 regression extrapolation
2 - use rounded 1891 regression extrapolation
3 - use unrounded 1901 regression extrapolation
4 - use rounded 1901 regression extrapolation
5 - use 1891 ratio extrapolation
6 - use 1901 ratio extrapolation
7 - use 1881 actual

TypeAlloc81
If newRecID recoded from worker (0) to entrepreneur (1) by imposition of Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers)

RReadyType
Final entrepreneur code

Type
Type as extracted from I-CeM

SubOccode
882 Sub-Occodes of economic activities

3.1.3.2. E 81

newRecID
Rec ID for 1881

E_corr81_91_pr_rou
prediction of being an employer (0=own account / 1=employer) by 1891 rounded regression

E_81_01_pr_rou
prediction of being an employer (0=own accout / 1=employer) by 1901 rounded regression

shortE
Supply of employers by 1 (1891 rounded regressions) or 2 (1901 rounded regressions)

RandAllEnt81
If newRecID has been randomly allocated from/to employer or own account

RandAllOAtoE81
If newRecID has been randomly allocated from own account to employer

RandAllEtoOA81
If newRecID has been randomly allocated from employer to own account

decision
Which extrapolation has been decided: 1.Use 1891 unrounded type extrapolation 2.Use 1891 rounded type extrapolation 3.Use 1901 unrounded type extrapolation 4.Use 1901 rounded type extrapolation 5.Use 1891 ratio extrapolation

TypeAllocE_81
If newRecID recoded from own account (0) to employer (1) by imposition of Type (Type 1 means employer)

ReadyTypeE_81
Final employer code

TypeE_81
Type as extracted from I-CeM

SubOccode
882 Sub-Occodes of economic activities

3.1.3.3. FARMER 81

TypeLogit_025
Final allocation of types for farmers after applying a reallocation of 5s to 1s with a logit farming model using a probability of 0.25 (odds ratio of 3)

3.1.3.4.

RECONSTRUCTION_NewEmployCodE variable New Employ Code using farmer reconstruction (non-optimised farmers' model)

3.1.4.

FINAL IMPORTANT VARIABLES (ALL YEARS = 51, 61, 81)

NewEmployCode (or variable NewEmployCode_Recon_NUM)
FINAL ALLOCATION OF THE RECONSTRUCTION = 1 worker, = 2 employer, = 3 own account

Employer_sample
0 = workers / 1 = Entrepreneurs: employer+own accounts

Farmer_sample
0 = non farmers/ 1 = Farmers

3.1.5. Others

-1 workers only
8 query? (too small to deal with)
888 SubOccodes of economic sector
9 non-economic

3.2. IND-method reconstruction variables (England and Wales).
(ALL YEARS = 51, 61, 81)

NewEmployCode_Recon_NUM
"FINAL ALLOCATION OF THE OTHER RECONSTRUCTION NUM = 1 worker, = 2 employer, = 3 own account"

SubOccode
882 Sub-Occodes of economic activities

coll_Ent
Average prediction of being an entrepreneur by SubOccode

sel_Ent
"If an individual is in a SubOccode with only workers (0), only entrepreneurs (1, non-observations) or a mix (2) before imposing the Types according to the 1891 logit regression"

pred_Ent
prediction of being an entrepreneur (0=worker / 1=entrepreneur) by 1891 logit regression

Worker_STRING
If an individual has been identified as a worker using his/her occupation string

Workers_only_SubOccode
If an individual belongs to a Worker-only SubOccode according to expert advice

Type
Type as extracted from I-CeM

Ent_rou_01
If an individual predicted probability of being an entrepreneur is higher than 0.1

Ent_rou_015
If an individual predicted probability of being an entrepreneur is higher than 0.15

Ent_rou_02
If an individual predicted probability of being an entrepreneur is higher than 0.2

Ent_rou_025
If an individual predicted probability of being an entrepreneur is higher than 0.25

Ent_rou_03
If an individual predicted probability of being an entrepreneur is higher than 0.3

Ent_rou_035
If an individual predicted probability of being an entrepreneur is higher than 0.35

Ent_rou_04
If an individual predicted probability of being an entrepreneur is higher than 0.4

Ent_rou_045
If an individual predicted probability of being an entrepreneur is higher than 0.45

Ent_rou_05
If an individual predicted probability of being an entrepreneur is higher than 0.5

Ent_rou_06
If an individual predicted probability of being an entrepreneur is higher than 0.6

Ent_rou_07
If an individual predicted probability of being an entrepreneur is higher than 0.7

Ent_rou_08
If an individual predicted probability of being an entrepreneur is higher than 0.8

Ent_r01aE_Wo

If an individual predicted probability of being an entrepreneur is higher than 0.1 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means
Entrepreneurs or 1) and Worker only (0) has been imposed

Ent_r015aE_Wo
If an individual predicted probability of being an entrepreneur is higher than 0.15 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed

Ent_r02aE_Wo
If an individual predicted probability of being an entrepreneur is higher than 0.2 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed

Ent_r025aE_Wo
If an individual predicted probability of being an entrepreneur is higher than 0.25 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed

Ent_r03aE_Wo
If an individual predicted probability of being an entrepreneur is higher than 0.3 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed

Ent_r035aE_Wo If an individual predicted probability of being an entrepreneur is higher than 0.35 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed

Ent_r04aE_Wo
If an individual predicted probability of being an entrepreneur is higher than 0.4 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed

Ent_r045aE_Wo
If an individual predicted probability of being an entrepreneur is higher than 0.45 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed
Ent_r05aE_Wo
If an individual predicted probability of being an entrepreneur is higher than 0.5 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed

Ent_r06aE_Wo
If an individual predicted probability of being an entrepreneur is higher than 0.6 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed

Ent_r07aE_Wo

If an individual predicted probability of being an entrepreneur is higher than 0.7 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed

Ent_r08aE_Wo
If an individual predicted probability of being an entrepreneur is higher than 0.8 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed

Decision
Which of the cut-offs has been decided by expert advice from the aggregate numbers of entrepreneurs and workers in each SubOccode

Ent_Decision
Whether the final allocation of an individual after the decision is an entrepreneur (1) or an worker (0)

coll_E
Average prediction of being an employer by SubOccode

sel_E_EW
"If individual is in a SubOccode with only own-accounts (0), only employers (1) or a mix (2) before imposing the Types according to the 1891 logit regression"

pred_E
prediction of being an employer (0=own account / 1=employer) by 1891 logit regression

E_rou_01
If an individual predicted probability of being an employer is higher than 0.1

E_rou_015
If an individual predicted probability of being an employer is higher than 0.15

E_rou_02
If an individual predicted probability of being an employer is higher than 0.2

E_rou_025
If an individual predicted probability of being an employer is higher than 0.25

E_rou_03
If an individual predicted probability of being an employer is higher than 0.3

E_rou_035
If an individual predicted probability of being an employer is higher than 0.35

E_rou_04
If an individual predicted probability of being an employer is higher than 0.4

E_rou_045
If an individual predicted probability of being an employer is higher than 0.45

E_rou_05
If an individual predicted probability of being an employer is higher than 0.5

E_rou_06
If an individual predicted probability of being an employer is higher than 0.6
E_rou_07
If an individual predicted probability of being an employer is higher than 0.7

E_rou_08
If an individual predicted probability of being an employer is higher than 0.8

E_r01aE
If an individual predicted probability of being an employer is higher than 0.1
OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

E_r015aE
If an individual predicted probability of being an employer is higher than 0.15 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

E_r02aE
If an individual predicted probability of being an employer is higher than 0.2 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

E_r025aE
If an individual predicted probability of being an employer is higher than 0.25 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

E_r03aE
If an individual predicted probability of being an employer is higher than 0.3 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed
E_r035aE
If an individual predicted probability of being an employer is higher than 0.35 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

E_r04aE
If an individual predicted probability of being an employer is higher than 0.4 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

E_r045aE
If an individual predicted probability of being an employer is higher than 0.45 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

E_r05aE
If an individual predicted probability of being an employer is higher than 0.5 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

E_r06aE
If an individual predicted probability of being an employer is higher than 0.6 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

E_r07aE If an individual predicted probability of being an employer is higher than 0.7 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

E_r08aE
If an individual predicted probability of being an employer is higher than 0.8 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

Decision_E
Which of the cut-offs has been decided by expert advice from the aggregate numbers of employers and own accounts in each SubOccode

E_Decision
Whether the final allocation of an individual after the decision is an employer (1) or an own account (0)

NewEmployCode_includingFarmers
Final allocation of the reconstruction IND before imposing the farm model (i.e. with the farm reconstruction IND of the general model)

NewEmployCode_Recon_IND
FINAL ALLOCATION OF THE RECONSTRUCTION IND = 1 worker, = 2 employer, = 3 own account"
comparison Whether the first reconstruction outcome is equal to (1) or different (0) from the reconstruction IND outcome

3.3. IND-method reconstruction variables (Scotland).
(ALL YEARS = 1851, 1861, 1871, 1881)

SubOccode
882 Sub-Occodes of economic activities

Occode
797 Occodes of economic activities

Ec_Ac
Economically active (=1 Worker or =10 Unknown Employ Code (W, E, or OA))

coll_Ent
Average prediction of being an entrepreneur by SubOccode

sel
If an individual is in a SubOccode with only workers (0), only entrepreneurs (1, non-observations) or a mix (2) before imposing the Types according to the 1891 logit regression

pred_Ent
prediction of being an entrepreneur (0=worker / 1=entrepreneur) by 1891 logit regression

ENT_afterExtracted_WorkersONLY
Entrepreneurs after extracted Workers only (only for 1881 census)

Type
Type as extracted from I-CeM

Ent_rou_01
If an individual predicted probability of being an entrepreneur is higher than 0.1

Ent_rou_015
If an individual predicted probability of being an entrepreneur is higher than 0.15

Ent_rou_02
If an individual predicted probability of being an entrepreneur is higher than 0.2

Ent_rou_025
If an individual predicted probability of being an entrepreneur is higher than 0.25

Ent_rou_03
If an individual predicted probability of being an entrepreneur is higher than 0.3

Ent_rou_035
If an individual predicted probability of being an entrepreneur is higher than 0.35

Ent_rou_04
If an individual predicted probability of being an entrepreneur is higher than 0.4

Ent_rou_045
If an individual predicted probability of being an entrepreneur is higher than 0.45

Ent_rou_05
If an individual predicted probability of being an entrepreneur is higher than 0.5

Ent_rou_06
If an individual predicted probability of being an entrepreneur is higher than 0.6

Ent_rou_07
If an individual predicted probability of being an entrepreneur is higher than 0.7

Ent_rou_08
If an individual predicted probability of being an entrepreneur is higher than 0.8

Ent_r01aE_Wo
If an individual predicted probability of being an entrepreneur is higher than 0.1 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed

Ent_r015aE_Wo
If an individual predicted probability of being an entrepreneur is higher than 0.15 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed

Ent_r02aE_Wo
If an individual predicted probability of being an entrepreneur is higher than 0.2 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed

Ent_r025aE_Wo
If an individual predicted probability of being an entrepreneur is higher than 0.25 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed

Ent_r03aE_Wo
If an individual predicted probability of being an entrepreneur is higher than 0.3 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed

Ent_r035aE_Wo
If an individual predicted probability of being an entrepreneur is higher than 0.35 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed

Ent_r04aE_Wo
If an individual predicted probability of being an entrepreneur is higher than 0.4 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed

Ent_r045aE_Wo
If an individual predicted probability of being an entrepreneur is higher than 0.45 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed

Ent_r05aE_Wo
If an individual predicted probability of being an entrepreneur is higher than 0.5 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed

Ent_r06aE_Wo
If an individual predicted probability of being an entrepreneur is higher than 0.6 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed

Ent_r07aE_Wo
If an individual predicted probability of being an entrepreneur is higher than 0.7 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed

Ent_r08aE_Wo
If an individual predicted probability of being an entrepreneur is higher than 0.8 OR Type (Types 1,3,5 for non-farmers and Types 1,2,5 for farmers means Entrepreneurs or 1) and Worker only (0) has been imposed

Decision
Which of the cut-offs has been decided by expert advice from the aggregate numbers of entrepreneurs and workers in each SubOccode

Ent_Decision
Whether the final allocation of an individual after the decision is an entrepreneur (1) or an worker (0)

coll_E _mean
Average prediction of being an employer by SubOccode

sel_E
If individual is in a SubOccode with only own-accounts (0), only employers (1) or a mix (2) before imposing the Types according to the 1891 logit regression

pred_E
prediction of being an employer (0=own account / 1=employer) by 1891 logit regression

E_rou_01
If an individual predicted probability of being an employer is higher than 0.1

E_rou_015
If an individual predicted probability of being an employer is higher than 0.15

E_rou_02
If an individual predicted probability of being an employer is higher than 0.2

E_rou_025
If an individual predicted probability of being an employer is higher than 0.25

E_rou_03
If an individual predicted probability of being an employer is higher than 0.3

E_rou_035
If an individual predicted probability of being an employer is higher than 0.35

E_rou_04
If an individual predicted probability of being an employer is higher than 0.4

E_rou_045
If an individual predicted probability of being an employer is higher than 0.45

E_rou_05
If an individual predicted probability of being an employer is higher than 0.5

E_rou_06
If an individual predicted probability of being an employer is higher than 0.6

E_rou_07
If an individual predicted probability of being an employer is higher than 0.7

E_rou_08
If an individual predicted probability of being an employer is higher than 0.8

E_r01aE
If an individual predicted probability of being an employer is higher than 0.1 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

E_r015aE
If an individual predicted probability of being an employer is higher than 0.15 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

E_r02aE
If an individual predicted probability of being an employer is higher than 0.2 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

E_r025aE
If an individual predicted probability of being an employer is higher than 0.25 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

E_r03aE
If an individual predicted probability of being an employer is higher than 0.3 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

E_r035aE
If an individual predicted probability of being an employer is higher than 0.35 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

E_r04aE
If an individual predicted probability of being an employer is higher than 0.4 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

E_r045aE
If an individual predicted probability of being an employer is higher than 0.45 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

E_r05aE
If an individual predicted probability of being an employer is higher than 0.5 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

E_r06aE
If an individual predicted probability of being an employer is higher than 0.6 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

E_r07aE
If an individual predicted probability of being an employer is higher than 0.7 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

E_r08aE
If an individual predicted probability of being an employer is higher than 0.8 OR Type (Types 1,3 for non-farmers and Types 1,2 for farmers means employer or 1) has been imposed

Decision_E
Which of the cut-offs has been decided by expert advice from the aggregate numbers of employers and own accounts in each SubOccode

E_Decision
Whether the final allocation of an individual after the decision is an employer (1) or an own account (0)

NewEmployCode_before_FarmingM
FINAL ALLOCATION OF THE RECONSTRUCTION BEFORE FARMING MODEL = 1 worker, = 2 employer, = 3 own account

ACREAGE
Acreage of the farm

TOTAL
Total employees

Ees
If a farmer has non-zero employees (1) or does not have employees at all (0)

EPortfolio
Portfolio

Type5
If an farmer is Type 5 (1) or not (0)

TypeLogit_025
FINAL ALLOCATION OF TYPES AFTER FARM MODEL (1-6)

Farmer_sample
(1) A FARMER OR (0) NOT A FARMER

NewEmployCode (or variable NewEmployCode_Recon_IND)
FINAL ALLOCATION OF THE SCOTTISH RECONSTRUCTION = 1 worker, = 2 employer, = 3 own account

3.3.1.
(Only 1851)

TypeLogit_025MA
FINAL ALLOCATION OF TYPES AFTER FARM MODEL (1-6) and Professor Mike Anderson deletion of duplicates and New ParID allocations

4. Conclusion

This paper gives the details of the reconstruction method that estimates the employment status of the full population of economically active individuals as worker, employer and own account for Censuses 1851, 1861, and 1881 (England and Wales) and 1851, 1861, 1871, and 1881 (Scotland). The main scope of the paper is to give worked examples to explain practically how the reconstruction is performed and how to dig into its most technical details. At the same time, it provides a comprehensive list of all the variables available in the accompanying reconstruction database that records each of the decisions taken during the reconstruction process. This is not a theoretical discussion but a series of worked examples and useful definitions to show how the reconstruction process is implemented.

Acknowledgments:

This research has been supported by the ESRC project ES/M010953 **Drivers of Entrepreneurship and Small Businesses**, which included a special supplementation to include Scotland so that full aggregates of Britain were available in the database deposit of the *British Business Census of Entrepreneurs* (BBCE). Piloting of the research for 1881 draws from Leverhulme Trust grant RG66385: **The long-term evolution of Small and Medium-Sized Enterprises (SMEs).** An Isaac Newton Trust Grant, 18.40(g) **Business proprietor succession and firm size change 1851-1881,** provided additional valuable support to help track and code Scottish employer data**.**

The database used for 1891 and 1901-11 derives from K. Schürer, E. Higgs, A.M. Reid, E.M Garrett, *Integrated Census Microdata, 1851-1911, version V. 2 (I-CeM.2),* (2016) [data collection]. UK Data Service, SN: 7481, http://dx.doi.org/10.5255/UKDA-SN-7481-1; enhanced; E. Higgs, C. Jones, K. Schürer and A. Wilkinson, *Integrated Census Microdata (I-CeM) Guide, 2nd ed*. (Colchester: Department of History, University of Essex, 2015).

This paper derives from the support of Principal Investigator Prof. Robert J. Bennett and project team members Harry Smith and Carry van Lieshout, University of Cambridge, Department of Geography and Cambridge Group for the History of Population and Social Structure, and draws from WP 9, and WP 9.2.

References

Bennett, R. J., Montebruno, P., Smith, H., & van Lieshout, C. (2018a). *Reconstructing entrepreneur and business numbers for censuses 1851-81. Working paper 9*. Cambridge, UK. https://doi.org/https://doi.org/10.17863/CAM.37738

Bennett, R. J., Smith, H., & Montebruno, P. (2018b). The Population of Non-corporate Business Proprietors in England and Wales 1891–1911. *Business History*. https://doi.org/10.1080/00076791.2018.1534959

Bennett, R. J., Montebruno, P., Smith, H., & van Lieshout, C. (2019a). *Reconstructing proprietor numbers for censuses 1851-81: Extension and alternative Working paper 9.2*. Cambridge, UK.

Bennett, R. J., Smith, H., van Lieshout, C., Montebruno, P., & Newton, G. (2019b). British Business Census of Entrepreneurs 1851-1911 (BBCE), UK Data Archive data deposit. *Forthcoming*.

Higgs, Edward and Schürer, Kevin (University of Essex) (2014) *The Integrated Census Microdata (I-CeM)* UKDA, SN-7481; K. Schürer, E. Higgs, A.M. Reid, E.M Garrett, *Integrated Census Microdata, 1851-1911, version V. 2 (I-CeM.2)*, (2016) [data collection] UK Data Service SN: 7481.

Montebruno, P., Bennett, R.J., van Lieshout, C., Smith, H. and Satchell, M., 2019. Shifts in agrarian entrepreneurship in mid-Victorian England and Wales. *Agricultural History Review*, 67, 71-108.

Other Working Papers:

Working paper series: ESRC project ES/M010953: *'Drivers of Entrepreneurship and Small Business',* University of Cambridge, Department of Geography and Cambridge Group for the History of Population and Social Structure.

WP 1: Bennett, Robert J., Smith Harry J., van Lieshout, Carry, and Newton, Gill (2017) *Drivers of Entrepreneurship and Small Businesses: Project overview and database design.* https://doi.org/10.17863/CAM.9508

WP 2: Bennett, Robert J., Smith Harry J. and van Lieshout, Carry (2017) *Employers and the self-employed in the censuses 1851-1911: The census as a source for identifying entrepreneurs, business numbers and size distribution.* https://doi.org/10.17863/CAM.9640

WP 3: van Lieshout, Carry, Bennett, Robert J., Smith, Harry J. and Newton, Gill (2017) *Identifying businesses and entrepreneurs in the Censuses 1851-1881*. https://doi.org/10.17863/CAM.9639

WP 4: Smith, Harry J., Bennett, Robert J., and van Lieshout, Carry (2017) *Extracting entrepreneurs from the Censuses, 1891-1911.* https://doi.org/10.17863/CAM.9638

WP 5: Bennett, Robert J., Smith Harry J., van Lieshout, Carry, and Newton, Gill (2017) *Business sectors, occupations and aggregations of census data 1851-1911*. https://doi.org/10.17863/CAM.9874

Data download of classification file: https://doi.org/10.17863/CAM.9874

WP 6: Smith, Harry J. and Bennett, Robert J. (2017) *Urban-Rural Classification using Census data, 1851-1911*. https://doi.org/10.17863/CAM.15763

WP 7: Smith, Harry, Bennett, Robert J., and Radicic, Dragana (2017) *Classification of towns in 1891 using factor analysis.* https://doi.org/10.17863/CAM.15767

WP 8: Bennett, Robert J., Smith, Harry, and Radicic, Dragana (2017) *Classification of occupations for economically active: Factor analysis of Registration Sub-Districts (RSDs) in 1891*. https://doi.org/10.17863/CAM.15764

WP 9: Bennett, Robert, J., Montebruno, Piero, Smith, Harry, and van Lieshout, Carry (2018) *Reconstructing entrepreneurship and business numbers for censuses 1851-81*. https://doi.org/10.17863/CAM.37738

WP 9.2: Bennett, Robert, J., Montebruno, Piero, Smith, Harry, and van Lieshout, Carry (2019) *Reconstructing business proprietor responses for censuses 1851-81: a tailored logit cut-off method.*

WP 10: Bennett, Robert, J., Smith, Harry and Radicic, Dragana (2018) *Classification of environments of entrepreneurship: Factor analysis of Registration Sub-Districts (RSDs) in 1891*. https://doi.org/10.17863/CAM.26386

WP 11: Montebruno, Piero (2018) *Adjustment Weights 1891-1911: Weights to adjust entrepreneur numbers for non-response and misallocation bias in Censuses 1891-1911*. https://doi.org/10.17863/CAM.26378

Adjustment weights: https://doi.org/10.17863/CAM.26376

WP 12: van Lieshout, Carry, Day, Joseph, Montebruno, Piero and Bennett Robert J. (2018) *Extraction of data on Entrepreneurs from the 1871 Census to supplement I-CeM.* https://doi.org/10.17863/CAM.27488

WP 13: van Lieshout, Carry, Bennett, Robert J. and Smith Harry (2019) *Extracted data on employers and farmers compared with published tables in the Census General Reports, 1851-1881*. https://doi.org/10.17863/CAM.37165

WP 14: van Lieshout, Carry, Bennett Robert J. and Montebruno, Piero (2019) *Company Directors: Directory and Census record linkage.* https://doi.org/10.17863/CAM.37166

WP 15: Bennett, Robert, J., Montebruno, Piero, Smith, Harry and van Lieshout, Carry (2019) *Entrepreneurial discrete choice: Modelling decisions between self-employment, employer and worker status.* https://doi.org/10.17863/CAM.37312

WP 16: Satchell, M., Bennett, Robert J., Bogart, D. and Shaw-Taylor, L. (2019) Constructing Parish-level Data and RSD-level Data on Transport Infrastructure in England and Wales 1851-1911. https://doi.org/10.17863/CAM.37313

WP 17: Satchell, M. and Bennett, Robert J. (2019) *Building a 1911 Historical Land Capacity GIS*. https://doi.org/10.17863/CAM.42285

WP 18: Bennett, Robert, J., Smith, Harry, van Lieshout, Carry and Montebruno, Piero (2019) *Identification of business partnerships in the British population censuses 1851-1911 for BBCE*.

WP 19: Montebruno, Piero (2019) *Datasets and guide: downloads for reconstructing British census responses 1851-1881 for the BBCE*.

WP 20: Smith, Harry, van Lieshout, Carry, Montebruno, Piero and Bennett, Robert, J. (2019) *Preparing Scottish census data in I-CeM for the British Business Census of Entrepreneurs (BBCE)*.

Full list of all current Working Papers available at:
http://www.geog.cam.ac.uk/research/projects/driversofentrepreneurship/